RESEARCH PAPER

Validating New Tuberculosis Computational Models with Public Whole Cell Screening Aerobic Activity Datasets

Sean Ekins • Joel S. Freundlich

ABSTRACT

Purpose The search for small molecules with activity against *Mycobacterium tuberculosis* (Mtb) increasingly uses high through-put screening and computational methods. Several public datasets from the Collaborative Drug Discovery Tuberculosis (CDD TB) database have been evaluated with cheminformatics approaches to validate their utility and suggest compounds for testing.

Methods Previously reported Bayesian classification models were used to predict a set of 283 Novartis compounds tested against Mtb (containing aerobic and anaerobic hits) and to search FDA approved drugs. The Novartis compounds were also filtered with computational SMARTS alerts to identify potentially undesirable substructures.

Results Using the Novartis compounds as a test set for the Bayesian models demonstrated a >4.0-fold enrichment over random screening for finding aerobic hits not in the computational models ($N = 34$). A 10-fold enrichment was observed for finding Mtb active compounds in the FDA drugs database. 85.9% of the Novartis compounds failed the Abbott SMARTS alerts, a value substantially higher than for known TB drugs. Higher levels of failures of SMARTS filters from different groups also correlate with the number of Lipinski violations.

Conclusions These computational approaches may assist in finding desirable leads for Tuberculosis drug discovery.

KEY WORDS Bayesian models · Collaborative Drug Discovery · function class fingerprints · mycobacterium tuberculosis · Quantitative Structure Activity Relationship · tuberculosis · tuberculosis database

INTRODUCTION

Mycobacterium tuberculosis (Mtb), the causative agent of tuberculosis (TB), infects approximately one third of the world's population, and 1.7–1.8 million people die from this disease annually [1]. Agents that are active against Mtb are urgently needed to combat this global epidemic that is heavily influenced by resistance to the available regimen of drugs, lengthy treatment, and co-infection with HIV.

S. Ekins
Collaborative Drug Discovery
1633 Bayshore Highway, Suite 342
Burlingame, California 94010, USA

S. Ekins (✉)
Collaborations in Chemistry
601 Runnymede Avenue
Jenkintown, Pennsylvania 19046, USA
e-mail: sekins@collaborativedrug.com
e-mail: ekinssean@yahoo.com

S. Ekins
Department of Pharmaceutical Sciences, University of Maryland
Baltimore, Maryland, USA

S. Ekins
Department of Pharmacology, Robert Wood Johnson Medical School
University of Medicine & Dentistry of New Jersey
Piscataway, New Jersey 08854, USA

J. S. Freundlich
Department of Biochemistry and Biophysics, Texas A&M University
College Station, Texas 77843, USA

Several recent reviews have summarized (2) and addressed the shortage of compounds in the research and development pipeline for this disease (3). This pipeline is overdue in delivering a drug, as there has not been a new approved treatment for tuberculosis in over 40 years. Recently, many phenotypic screening efforts have searched for compounds that inhibit the growth of Mtb (3). These compounds broadly sample libraries of small molecules, yet such studies have rarely used past knowledge of Mtb active compounds to focus screening. Leveraging such knowledge to produce computational models or rules for the virtual screening of compound libraries as a complement to high-throughput screening *in vitro* could improve the efficiency of screening (4). Computational and statistical analyses can also be utilized to provide insights into small molecule physiochemical properties which may be important for activity against Mtb. Additionally, several ligand-based (5) or protein-based (6) studies have discussed filters that may be implemented either before or after a biological screen in order to identify molecules with an optimal set of physiochemical properties (e.g. hit-like, lead-like, or drug-like (7)) to further streamline the discovery of novel antituberculars.

We have previously used Bayesian methods (5) with molecular function class fingerprints of maximum diameter 6 (FCFP_6) (8) to identify substructures that are important in recent tuberculosis screening datasets (9). We have also extended this approach and validated the models with a set of 102,000 compounds from the same laboratory containing 1702 molecules with ≥90% inhibition at 10 μM, representing a hit rate of 1.66% (10). We were able to demonstrate 10-fold enrichments in finding active compounds in the top ranked 600 molecules (10).

Critically, we have sought to utilize cheminformatics to not only predict small molecules with activity against Mtb, but to leverage methods to select for chemical entities, either as library inputs or screening hits, with a desired set of physiochemical properties (10–12). Filters can enable removal or flagging of undesirable molecules (thiol traps and redox-active compounds, epoxides, anhydrides, and Michael acceptors that can covalently modify a cysteine moiety in a surrogate protein (13–15)), false positives and frequent hitters (16). For example, we have previously compared the filtering of malaria hits and datasets screened against Mtb, and three antimalarial datasets had very high failures with the Abbott Alerts (11), while a similar pattern was seen for Mtb (> 81% failure) *versus* known Mtb drugs (> 50% failure) (10).

In the current study, we have used two new independent test sets for our Bayesian models of Mtb whole cell efficacy in order to further demonstrate the predictive value of these statistical models in Mtb drug discovery. One data set consists of 283 compounds screened against cultured Mtb from Novartis (details available at http://www.collaborativedrug.

com/register), and we have used it to successfully identify the 34 compounds with MIC values <10 μM. After removing compounds already in the Bayesian model, the test set size was reduced to 248 molecules. This complete Novartis dataset was also filtered using SMARTS filters (11,12) and compared to our previous analysis of other libraries screened against Mtb. Additionally, we have used our Bayesian models for Mtb to screen natural products and FDA approved drugs to further validate our approach against published data (17) in order to suggest new compounds that may be useful to test *in vitro*. The contributions of this study are 1) further validation of Bayesian models of Mtb whole cell efficacy to deliver actives with a high degree of enrichment over random sampling and 2) demonstration of the utility of SMARTS filters to suggest potential liabilities and for refinement of either the chemical library input or hit set output for a screen. These along with similar cheminformatics methods (18,19) could help to identify new molecules that could become useful probes of Mtb essential biology or inspire novel antitubercular leads or drugs.

MATERIALS AND METHODS

CDD Database

The development of the CDD TB database (Collaborative Drug Discovery Inc. Burlingame, CA) has been previously described (20). Screening datasets were collected and uploaded in CDD TB from sdf files and mapped to custom protocols (21). We have also used a separate database of antibiotics (N=163) obtained from the Microsource US Drugs database (N=1039) as well as a large dataset of approved drugs (N=2815 were used for predictions with the Bayesian Models) from Professor David Sullivan (Johns Hopkins University). A subset of FDA approved drugs (N=2804) was further grouped into those with 0, 1, 2, 3 or 4 Lipinski violations (22). A set of 283 Novartis compounds screened against Mtb under aerobic and anaerobic conditions was recently kindly provided and made available in the CDD TB database.

Descriptors

The Novartis compounds and, in particular, those that were aerobic or anaerobic hits were compared to MLSMR and TAACF-NIAID-CB2 hits (10,20) using simple and readily interpretable calculated molecular properties including logP, number of hydrogen bond donor, number of hydrogen bond acceptor, Lipinski Rule of Five alerts, polar surface area, molecular weight, rotatable bonds, and atom counts, which were computed using the Marvin plug-in (ChemAxon, Budapest, Hungary) within the CDD database (10,20).

Machine Learning with 2D Descriptors

We have previously described the Bayesian classifier modeling method in detail [23]. Two Laplacian-corrected Bayesian classifier models were generated using Discovery Studio 2.1 or 2.5.5 (Accelrys, San Diego, CA) [10,20] as described previously for the MLSMR 220,463 molecules (4096 active) [9] and the dose-response data using 2273 molecules (475 active). These models used molecular function class fingerprints of maximum diameter 6 (FCFP_6) [8] and the following interpretable descriptors: AlogP [24], molecular weight, number of rotatable bonds, number of rings, number of aromatic rings, number of hydrogen bond acceptors, number of hydrogen bond donors, and molecular fractional polar surface area. It has been described elsewhere that pre-selection of descriptors is not required with Bayesian methods, as only descriptors correlating with activity are used and other descriptors discarded. Hence, there is less of an over-fitting issue [25]. These two Laplacian-corrected Bayesian classifier models for Mtb have been previously validated with several external test sets [9] including the TAACF- NIAID-CB2 dataset of 102,634 molecules [10].

The Novartis compounds as well as the FDA approved drug database were downloaded from the CDD databases. These datasets as well as a set of 800 natural products (from the Microsource natural products database, http://www. msdiscovery.com/) were also scored with the Bayesian models. Any molecules that were identical (Tanimoto similarity using MDL fingerpint keys = 1) to those in the training set were removed to afford 248 Novartis compounds, 663 natural product molecules and 2108 FDA approved drugs. As a comparison of test sets, the mean maximal Tanimoto similarity for test set compounds ± SD was calculated using the MDL fingerprints in Discovery Studio.

Predictions for the datasets were calculated from the input sdf file using the "calculate molecular properties" protocol to distinguish between compounds that are active against Mtb and those that are inactive under aerobic conditions. The data were then ranked, and the prioritization of hits was then used to create Receiver Operator Characteristic (ROC) graphs in Microsoft Excel 2003.

A new Bayesian model was generated using the 283 Novartis compounds (42 with aerobic activity MIC < 10 μM and classified as active) as a training set. Molecular function class fingerprints of maximum diameter 6 (FCFP_6) [8], AlogP, molecular weight, number of rotatable bonds, number of rings, number of aromatic rings, number of hydrogen bond acceptors, number of hydrogen bond donors, and molecular fractional polar surface area were used as the descriptors with the "calculate molecular properties" protocol. This model was further validated using leave-one-out cross-validation. Each sample was left out one at a time, and a model built using the results of the samples was used to predict the left-out sample. Once all the samples had predictions, a ROC plot was generated, and the cross-validated ROC area under the curve (XV ROC AUC) was calculated [23]. Other statistics were also generated as previously described elsewhere [20,23]. The model was additionally evaluated by leaving out 50% of the data and rebuilding the model 100 times using a custom protocol in Discovery Studio (available from the author on request) for validation, in order to generate the ROC AUC [23]. We also used various previously described datasets as external test sets for this model.

SMARTS Filters

The Abbott ALARM [13], Glaxo [26] and Pfizer LINT SMARTS (also called the Blake filters [27]) filter calculations were performed through the SMARTS filter web application, kindly provided by the Division of Biocomputing, Department of Biochemistry and Molecular Biology, University of New Mexico, Albuquerque, NM, (http://pasilla.health.unm.edu/tomcat/biocomp/smartsfilter). This software identifies the number of compounds that pass or fail any of the filters implemented. Each filter was evaluated individually with each set of compounds.

The SMARTS filters in Discovery Studio 2.5.5 were also used as previously described [10]. Each filter has a minimum and maximum number of times that it is allowed to map. A molecule must match this filter or it will be noted as failing the filter.

RESULTS

Differentiating Aerobic from Anaerobic Mtb Hits

With the Novartis Mtb screening data, the anaerobic compounds in several cases showed statistically different and higher mean descriptor property values compared with the aerobic hits (e.g. molecular weight, logP, hydrogen bond donor, hydrogen bond acceptor, polar surface area and rotatable bond number, Table I). The mean molecular properties for the Novartis compounds are in a similar range to the MLSMR and TAACF-NIAID CB2 hits (Table I).

Bayesian Model Development and Validation

Laplacian-corrected Bayesian classifier models are computationally fast and have been used widely for several drug discovery applications in recent years, including with Mtb [5]. In this study, we first used the Novartis aerobic assay

Table I Novartis Hits Mean (SD) Molecular Descriptor Property Values Compared with MLSMR and TAACF Dataset Hits Described Below (10,20). Mean Aerobic and Anaerobic Hits from the Novartis Dataset were Compared with a Two Sided t-test, [a]$p < 0.05$, [b]$p < 0.0001$.

Dataset	Molecular weight	logP	Number of hydrogen bond donors	Number of hydrogen bond acceptors	Lipinski Rule of Five alerts	Atom count	Polar surface area	Rotatable bond number
Novartis compounds (N = 283)	331.23 (78.06)	3.30 (1.44)	1.24 (0.98)	4.18 (1.97)	0.13 (0.40)	39.12 (10.51)	69.75 (34.29)	3.72 (1.99)
Novartis Aerobic hits < 10 μM (N = 42)	312.23 (59.79)	2.92 (1.33)	0.88 (0.77)	3.26 (1.72)	0.07 (0.26)	38.09 (10.54)	49.41 (23.75)	3.00 (1.85)
Novartis Anaerobic hits (N = 241)	334.54 (80.47)[a]	3.36 (1.45)[a]	1.31 (1.00)[a]	4.34 (1.97) [a]	0.14 (0.42)	39.30 (10.52)	73.29 (34.65)[b]	3.84 (1.99)[a]
MLSMR Active ≥ 90% inhibition at 10 μM (N = 4096)	357.10 (84.70)	3.58 (1.39)	1.16 (0.93)	4.89 (1.94)	0.20 (0.48)	42.99 (12.70)	83.46 (34.31)	4.85 (2.43)
TAACF-NIAID CB2 Active ≥ 90% inhibition at 10 μM (N = 1702)	349.58 (63.82)	4.04 (1.02)	0.98 (0.84)	4.18 (1.66)	0.19 (0.40)	41.88 (9.44)	70.28 (29.55)	4.76 (1.99)

hits as a test set for the previously described Bayesian models (10,20) to provide further validation of the approach using the published data from a different group. The mean maximal Tanimoto similarity was 0.67 ± 0.13 for the dose response model and 0.48 ± 0.12 for the single point model (using the MDL fingerprints). ROC plots were generated for both the Southern Research Institute (SRI) single point model and the SRI dose-response model (Fig. 1). For the Novartis data, over 35% of the total hits are found with the dose-response model in the top 8% of molecules (enrichment > 4-fold). Interestingly the single point model performed initially only slightly better than random, then degraded rapidly to random (Fig. 1).
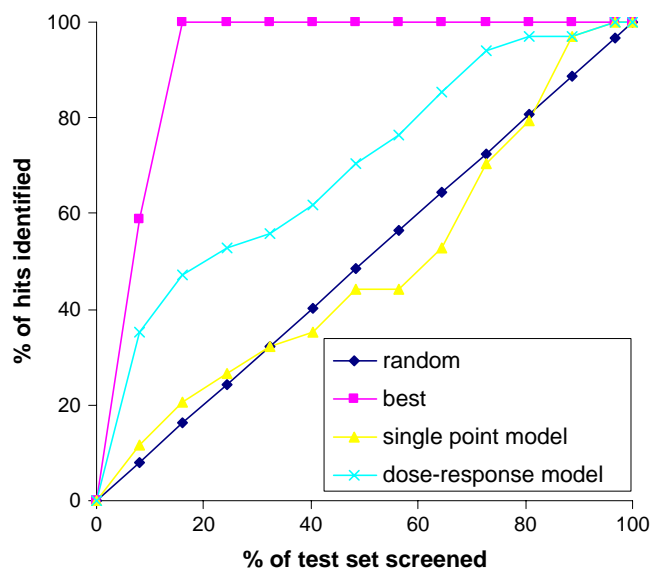

Fig. 1 Receiver operator characteristic for the Novartis aerobic Mtb hits (N = 34) used as a test set (N = 248) for the two previously published Bayesian models (20).
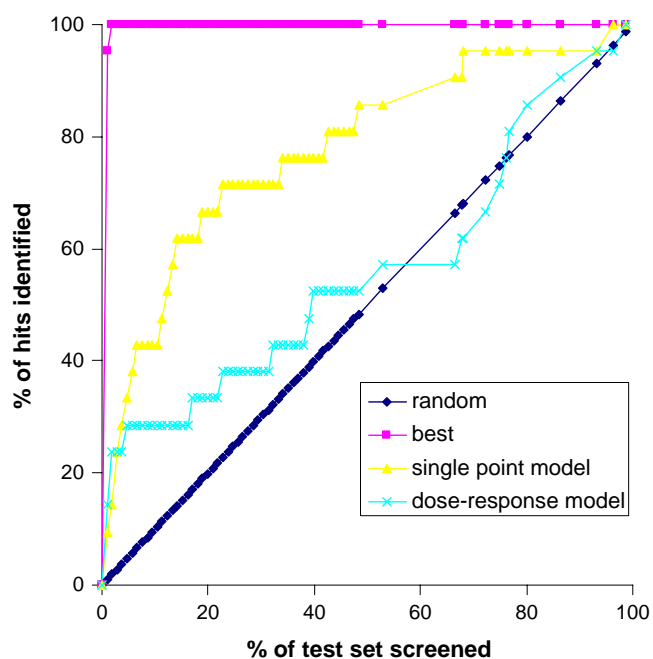
Recently, a set of 1514 known drugs (both FDA and foreign approved) was screened against Mtb and the MIC values determined using the Alamar blue susceptibility assay (17). Compounds were identified that had known antitubercular activity (N = 20), and novel hits (N = 18) were found. We have used 21 of these actives (Table II) as a second test set seeded in a larger set of 2108 FDA approved molecules downloaded from the CDD database. The mean maximal Tanimoto similarity of these compounds was 0.22 ± 0.08 for the dose-response model and 0.59 ± 0.17 for the single point model. We were able to show that both models initially had ~ 10-fold enrichments (Fig. 2). The single point model identified >60% of the active compounds (Table II) in the top 14% of all FDA approved compounds, while the dose response model was initially comparable but rapidly degraded in prediction quality (Fig. 2).

In addition to utilizing the Novartis data as a test set for the SRI models, we have used it to create a new Bayesian model. The Novartis Bayesian TB model had leave-one-out-cross-validation ROC statistics (0.84, Table III) that were also stable after leaving out 50% of the data 100 times (Table IV). The latter validation method also showed good concordance and specificity statistics but low sensitivity (Table IV). Figure 3 shows molecular substructure features important for discriminating between active and inactive compounds in the Novartis Bayesian TB model (the training set). It is intriguing to note that the imidazole fragment, present in the Phase II investigational drugs PA-824 (28) and OPC-67683 (29) is quite prevalent amongst these actives, as was the amide linker. Perturbation of the arrangement of ring nitrogens from 1,3- to 1,2 afforded pyrazole- and pyrazolone-derived molecules that were inactive in most cases. This Novartis model was used to predict active compounds in the SRI dose-response dataset (473 actives in 2267 compounds, mean Tanimoto

Table II Rank Ordering of FDA Approved Drugs with the Bayesian Mtb Models and Previously Published Literature MIC Values (17)

Drugs	MIC vs. Mtb (μM) (17)	Single point Bayesian score (rank)	Dose-response Bayesian score (rank)
Lomefloxacin	5	31.52 (13)	21.70 (5)
Pyrvinium Pamoate	0.31	29.55 (15)	−3.62 (1686)
Triple Dye (Brilliant Green, Gentian Violet, Proflavine Hemisulfate)	5	18.61 (34)	−4.76 (1822)
Pyrithione	2.5	14.62 (47)	−1.00 (1117)
Oxiconazole nitrate	>5	13.89 (51)	3.35 (95)
Minocycline	0.31	9.83 (79)	15.09 (10)
Aconiazide	0.63	8.33 (98)	−0.31 (823)
Methylene blue	>5	7.92 (108)	0.89 (356)
Doxycycline Hyclate	0.16	6.11 (129)	9.65 (17)
Tetracycline	1.25	2.10 (238)	7.60 (26)
Cetalkonium Chloride	5	1.59 (258)	0.55 (462)
Doxycycline	0.63	0.93 (280)	8.00 (25)
Primaquine	5	0.80 (285)	−3.26 (1617)
Chlorhexidine	5	−0.93 (390)	−3.22 (1608)
Methylbenzethonium Chloride	5	0.037 (475)	−2.76 (1523)
Bismuth subnitrate	2.5	−2.76 (706)	−0.30 (818)
Thonzonium	5	−3.77 (899)	−3.07 (1577)
Hexadimethrine Bromide	5	−4.50 (1010)	0.01 (665)
Demecarium	>5	−7.75 (1399)	−2.32 (1429)
Sulfathiazole	>5	−8.09 (1433)	−12.7 (2082)
Cefmenoxime	2.5	−32.95 (2031)	−6.88 (1963)

similarity 0.39±0.09). The enrichment in the top-ranked 100 compounds was approximately 2-fold, and this rapidly declined to random (data not shown). Similarly, we used

the FDA dataset (21 actives in 2815 compounds, mean Tanimoto similarity 0.27±0.09) such that there were no compounds in common with the training set. The Novartis model performed essentially random on this dataset (data not shown). In all test case examples described above it appears that better results are obtained with the training set that possesses the higher mean maximal Tanimoto similarity. With the Novartis data, these mean maximal Tanimoto similarity values are very low.

SMARTS Filters

The Novartis Mtb compounds were examined with several well-known filters for offending compounds or "swill" (30), which we have used in recent studies (10–12). 85.9% failed the Abbott Alarm filters, 47.7% failed the Pfizer LINT filters, 7.1% failed the GSK filters, and 37.4% failed the

Fig. 2 Receiver operator characteristic for the FDA approved Mtb hits (N = 21) used as a test set (N = 2108) for the two previously published Bayesian models (20).

Table III Cross Validated Results for Bayesian Model Building (FN = False Negative, FP = False Positive, LOO = Leave One Out, LO = Leave Out%, ROC = Receiver Operator Characteristic, TN = True Negative, TP = True Positive)

Dataset (number of molecules)	XV ROC AUC	Best Split	TP/FN FP/TN	Number of actives
Novartis (N = 283)	0.84	1.115	28/14 20/221	42

Table IV Mean (SD) Leave Out 50% x 100 Cross Validation of Bayesian Model (ROC = Receiver Operator Characteristic)

Dateset (number of molecules)	External ROC Score	Internal ROC Score	Concordance	Specificity	Sensitivity
Novartis (283)	0.81 (0)	0.82 (0)	80.61 (9.0)	84.19 (12.4)	59.91 (15.0)

Accelrys filters (Table V). The SMARTS filtering data (Table V) are consistent with what was seen in the analysis of the other TB datasets, FDA approved drugs, GSK and Novartis malaria data, etc. (12). The high level of failures with the Abbott ALARM filter is a potential concern and has been observed previously (10–12). Interestingly, while a natural products database has a lower failure rate for the Abbott ALARM and Pfizer LINT filters, the GSK and Accelrys filters had a higher failure rate compared with the Novartis compounds.
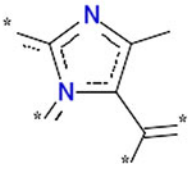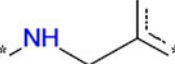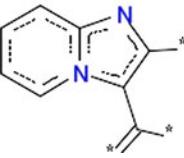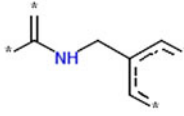
a
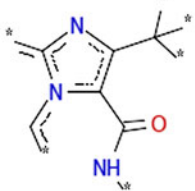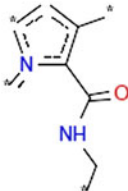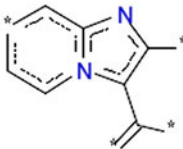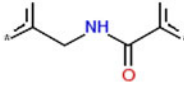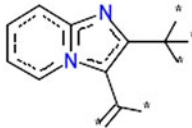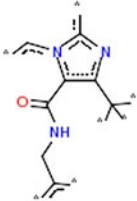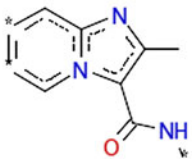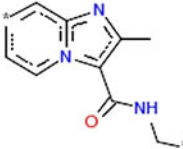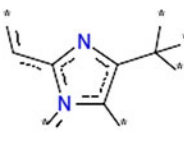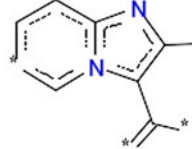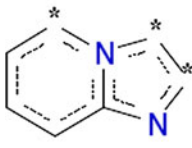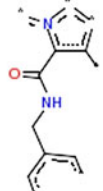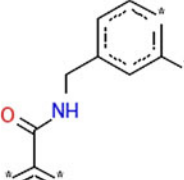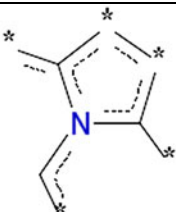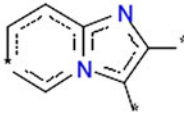
G1: -803711802
12 out of 12 good
Bayesian Score: 1.536

G2: -904785030
12 out of 12 good
Bayesian Score: 1.536

G3: 144183755
12 out of 13 good
Bayesian Score: 1.484

G4: -1350693998
10 out of 10 good
Bayesian Score: 1.483

G5: 1363933490
10 out of 10 good
Bayesian Score: 1.483

G6: 737835132
10 out of 10 good
Bayesian Score: 1.483

G7: -641378730
9 out of 9 good
Bayesian Score: 1.449

G8: 1592795550
9 out of 9 good
Bayesian Score: 1.449

G9: 444547831
18 out of 25 good
Bayesian Score: 1.388

G10: -1644617620
9 out of 10 good
Bayesian Score: 1.387

G11: 624859675
10 out of 12 good
Bayesian Score: 1.369

G12: 845414188
7 out of 7 good
Bayesian Score: 1.362

G13: -1057125929
7 out of 7 good
Bayesian Score: 1.362

G14: 973813398
7 out of 7 good
Bayesian Score: 1.362

G15: -203115083
14 out of 20 good
Bayesian Score: 1.323

G16: -444825055
14 out of 20 good
Bayesian Score: 1.323

G17: 27258599
9 out of 12 good
Bayesian Score: 1.274

G18: 907096426
10 out of 14 good
Bayesian Score: 1.268

G19: -670163659
5 out of 5 good
Bayesian Score: 1.233

G20: -2102302457
7 out of 9 good
Bayesian Score: 1.226

Fig. 3 Bayesian model for Novartis whole cell aerobic data. (**a**). Simple descriptors with FCFP_6: features important for Actives, (**b**). Simple descriptors with FCFP_6: features important for Inactives. * = any atom.
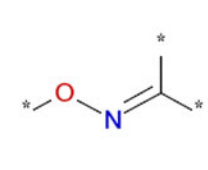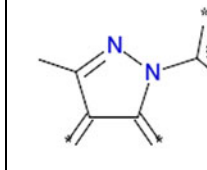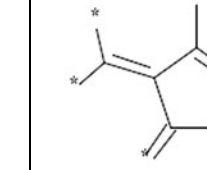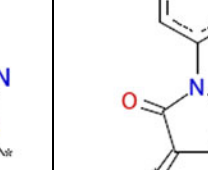
b
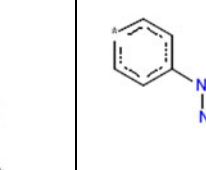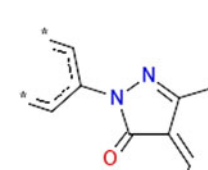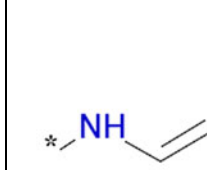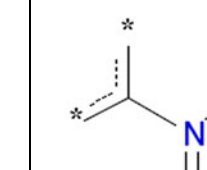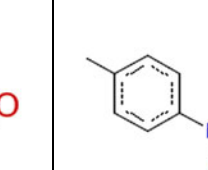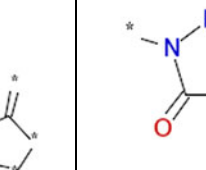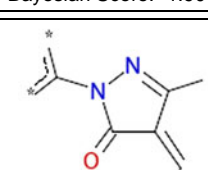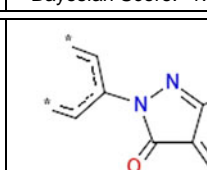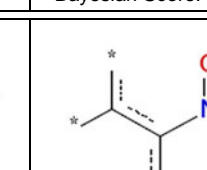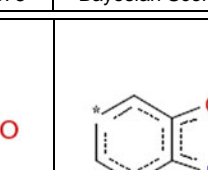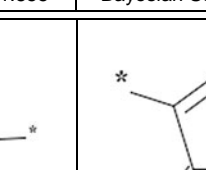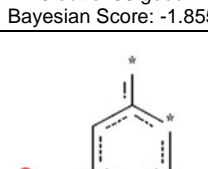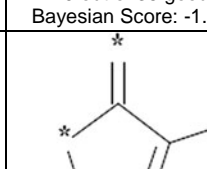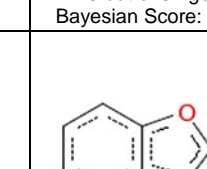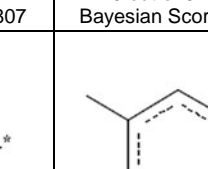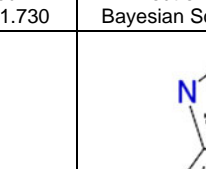
B1: 580960234
0 out of 54 good
Bayesian Score: -2.207

B2: -1764774431
0 out of 49 good
Bayesian Score: -2.121

B3: -760758590
0 out of 47 good
Bayesian Score: -2.084

B4: -236917003
0 out of 40 good
Bayesian Score: -1.944

B5: -1491838316
0 out of 39 good
Bayesian Score: -1.923

B6: -842275882
0 out of 38 good
Bayesian Score: -1.901

B7: 451937097
0 out of 38 good
Bayesian Score: -1.901

B8: -262018472
0 out of 37 good
Bayesian Score: -1.878

B9: -1036166322
0 out of 36 good
Bayesian Score: -1.855

B10: -864950439
0 out of 36 good
Bayesian Score: -1.855

B11: -1317177205
0 out of 36 good
Bayesian Score: -1.855

B12: -947530708
0 out of 35 good
Bayesian Score: -1.831

B13: 2037274029
0 out of 34 good
Bayesian Score: -1.807

B14: 193705859
0 out of 31 good
Bayesian Score: -1.730

B15: 565998553
1 out of 67 good
Bayesian Score: -1.708

B16: 876242371
0 out of 30 good
Bayesian Score: -1.703

B17: 436886043
1 out of 63 good
Bayesian Score: -1.652

B18: -1781982802
0 out of 26 good
Bayesian Score: -1.588

B19: 1075826633
0 out of 25 good
Bayesian Score: -1.557

B20: 690481386
0 out of 23 good
Bayesian Score: -1.492

Fig. 3 (continued)

For comparison, we have evaluated the FDA approved drugs from CDD and looked at the subsets with 0 to 4 Lipinski violations (Supplemental Tables I and II). Molecules with zero Lipinski violations had the lowest levels of

Table V SMARTS Filtering Number of Failures (%) Using the SMARTS Filter Website. The Discovery Studio Software was also Used as a Comparison.

Dataset	Abbott ALARM	Pfizer LINT	GSK	Accelrys
Novartis (283)	243 (85.9)	135 (47.7)	20 (7.1)	91 (37.4)
TB drugs (13)	7 (53.8)	6 (46.1)	1 (7.7)	9 (69.2)
US Antibiotic drugs from Microsource US drugs database (163)	144 (88.3)	93 (57.0)	57 (35)	126 (77.3)
Microsource US FDA drugs (1041)	688 (66.1)	516 (49.6)	143 (13.7)	499 (47.9)
Johns Hopkins –All FDA drugs (N = 2693)	1442 (53.5)	1264 (46.9)	401 (14.9)	1357 (50.4)
Natural products from Microsource (800)	521 (65.1)	304 (38.0)	125 (15.6)	397 (49.6)

Fig. 4 (**a**) A plot of the percentage of SMARTS filter failures for compounds with different numbers of Lipinski violations. (**b**) A plot of percentage of FDA drugs and the different numbers of Lipinski violations.

SMARTS filter failures across all four filters. We show a correlation between the number of SMARTS filter failures and the number of Lipinski violations for all different types of rules sets (Fig. 4A). The Novartis compounds that 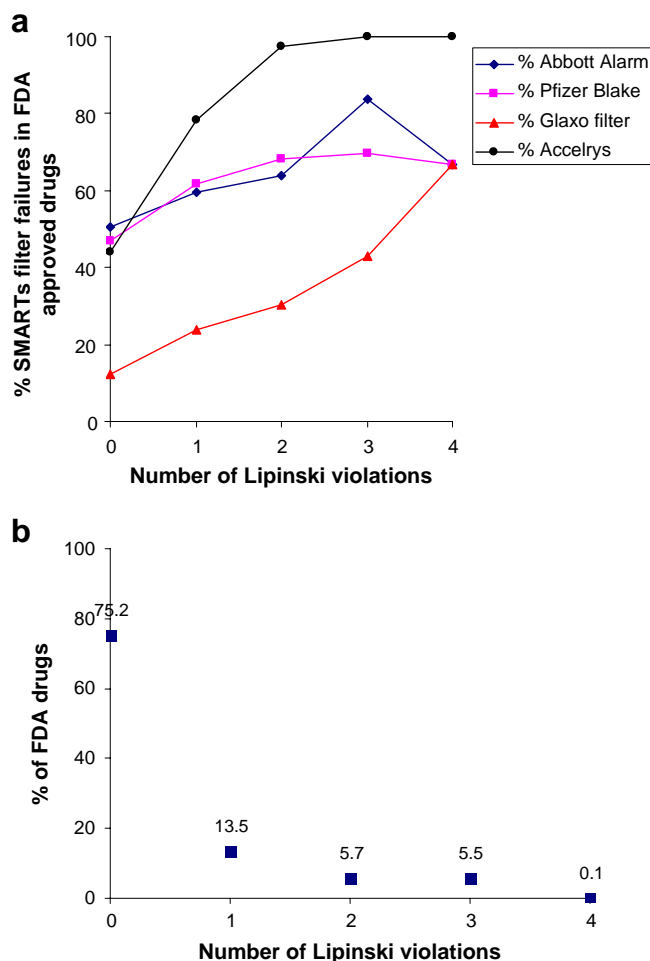are aerobic hits all have either 0 or 1 Lipinski violations, while there are a small number of aerobic inactives that have more violations (Supplemental Table III). When the Novartis dataset SMARTS filter failures are analyzed, there is little difference between the Abbott Alerts percentage failures in aerobic actives or non-actives, while the other filters are more discriminating, with a higher failure rate for non-actives *versus* actives (Supplemental Table IV).

DISCUSSION

Efforts to develop useful chemical probes of Mtb with the potential to inspire novel lead compounds have seen relatively limited application of cheminformatics methods that have been utilized in other therapeutic areas (5,19,31).

Although not the first to implement Bayesian analyses of Mtb data, we have previously used a far larger dataset (over 200,000 compounds) compared with <4000 compounds reported by others (5). We have also previously used the training sets (9) to identify many unique molecular fragments that are present in all actives, which could be useful in helping define the chemical attributes of an antitubercular compound to more effectively seed drug design efforts. These models suggest the value of continued learning in such computational methods (e.g. progressive increase in the size of a model as new data are added) and validation with different test sets. Previously, we have validated both Bayesian Mtb models with external compounds using the published NIAID, GVKbio and TAACF-NIAID CB2 datasets, which range from 2880 to over 102,000 compounds. The TAACF-NIAID CB2 data came from the same source (9,32) as the training sets used in the original models and represents an ideal scenario from modeling to limit any experimental variability. The largest test set also contained a more realistic percentage of hits, ~ 2%. In this example we showed in this case that Bayesian models could enhance the number of hits identified 10-fold over random high throughput screening. The caveat with these models is they are likely limited to predicting compound activity under the exact *in vitro* conditions used (9) and within the chemistry space of the model training sets (a limitation for all ligand-based computational methods). The single point Mtb data has much greater chemical space coverage than the dose-response data, as the dataset is 100 times larger.

With the availability of an additional Mtb test set provided by Novartis (the first made available from a major pharmaceutical company to our knowledge) we can now explore further how these Bayesian models (20) predict molecules from a single independent external laboratory. When we analyzed the mean molecular properties for the Novartis compounds containing both aerobic and anaerobic hits, in total they were in a similar range to the MLSMR and TAACF-NIAID CB2 hits described previously (Table I), which gave us some confidence that the compounds were not going to be dramatically different (e.g. large natural products). However, when we focused on just the 42 compounds with aerobic MIC values < 10 μM (so as to more closely represent the data from the original Bayesian model training sets), these property values were lower in all cases compared with the total dataset of aerobic and anaerobic data. The anaerobic compounds in several cases showed statistically different and higher mean descriptor property values compared with the aerobic hits (e.g. molecular weight, logP, hydrogen bond donor, hydrogen bond acceptor, polar surface area and rotatable bond number). This confirms our previous observations that such generally normally distributed properties across the datasets

can be used as potential ideal property range targets when looking at new sets of molecules. In this case, perhaps we can also discriminate between likely anaerobic *versus* aerobic hits using molecular properties. When the Novartis compounds were used as a test set (34 hits in 248 molecules, after removal of compounds already in the Bayesian model), we found a >4-fold initial enrichment over random screening with the dose-response Bayesian model [20]. The single point Bayesian model did not perform as well and also had a lower mean maximal Tanimoto similarity than the dose-response model, suggesting this measure may be a useful guide for model prediction quality. Although previously we have seen a 10-fold enrichment with a test set of over 100,000 compounds, in the current example we used less than 250 compounds.

A second recently generated external test set was also used, from a recent study by the Medical Research Council screening over 1000 FDA approved drugs against Mtb and presented 53 hits [17]. We have described similarities and differences in our predictions for 21 of these compounds (which were not part of the training sets of the formerly reported Bayesian models [20], Table II). Searching the FDA drugs for these Mtb hits with the Bayesian models [20] derived from data from SRI indicated a 10-fold enrichment. We also leveraged the Bayesian models [20] to rank a series of natural products as well as the FDA approved drugs. After removing those structures identical to those in the model, we arrived at some potential compounds of interest (Supplemental Fig. 1). The dose-response model demonstrated sertaconazole (antifungal), clofarabine (antineoplastic), tioconazole (antifungal) and amodiaquine (antimalarial) to be highly scored. The single point model demonstrated quinaldine blue (antineoplastic), atorvastatin (anti-hyperlipidemic) and montelukast (antiasthmatic) to be highly scored. In the natural products dataset, the single point Bayesian model [20] ranked daunorubicin and 4′-methoxychalcone with very high scores. The dose-response model ranked inosine and hieracin, iridin, harmane, and irigenol as high scoring. This suggests that these compounds should be screened *versus* Mtb, as the majority of these have not been previously reported to exhibit antitubercular efficacy (Supplemental Fig. 1). Following searching of the various TB datasets in the CDD TB database, harmane and daunorubicin were found to inhibit Mtb growth as judged in an Alamar blue-based assay (40 and 95% respectively) by the TB early phase drug discovery group (data kindly provided by Dr. Bernard Munos and available from CDD). Searching PubChem showed that sertaconazole (MIC_{90}=3.4 μg/mL *vs*. H37Rv) and daunorubicin (PubChem MIC=0.169 μg/mL *vs*. H37Rv) were active, while inosine (inactive against non-replicating, drug-tolerant Mtb) and harmane (not active in SRI screen) were not. Harmane represents an example showing weak activity

in the Alamar blue and no activity in whole cells, while the data for daunorubicin suggest it is active in both assays described above and tends to suggest some inter-laboratory variability in screening against Mtb. On the whole, these findings provide further prospective validation for the models.

The test set results suggest that the very large Bayesian models [20] generated with whole cell screening data from one laboratory (in this case SRI) could be used to reliably rank compounds screened and identified as Mtb hits by two independent groups.

Interestingly, the Bayesian dose response model performs better at retrieving the Novartis actives, while the converse is true for the FDA approved drugs tested against Mtb by the Medical Research Council [17]. This is in contrast to what was shown previously with the NIAID, GVK or TAACF-CB2 test datasets [10,20], which indicated both models, the dose response and single point, performed similarly. This represents a benefit of using multiple computational models, as one may perform better than the other depending on the similarity of the test molecules, as described earlier. Focusing on the model with the highest mean maximal Tanimoto similarity may also be justified. In general, when scoring compound libraries, it may therefore be possible to look at compounds ranked highly by both models (if the mean maximal Tanimoto similarity values are close) or form a consensus between both models, i.e. prioritizing those scoring highly in both models over those scoring well in just a single model.

A Bayesian model was also generated with the Novartis whole cell screening data using the aerobic active compounds. This appeared to perform well upon use of the more conservative leave out 50% 100-fold analysis (e.g. it was internally consistent). However, when tested with the SRI dose response and the FDA drug datasets, this model did not perform as well as those described earlier. This could be due to the limited coverage of the training set (283 *vs*. >2000 *vs*. >220,000 compounds, for the Novartis, SRI dose-response or SRI single point screening models, respectively), which is also reflected in the low mean maximal Tanimoto similarity values. In particular, this is seen when trying to predict the 21 actives in the FDA datasets, as the mean maximal Tanimoto similarity for these was < 0.2 alone; in addition, the mean similarity values for all the SRI data (0.39) and FDA data (0.27) are very low. An alternative explanation could also be the differences in the *in vitro* assays used to afford the various datasets. However, this appears unlikely based on the previously described results showing good enrichments using the SRI datasets with data from these other groups. The results point to the importance of using very large high throughput screening datasets for Bayesian model building that will provide good coverage of future chemical libraries

or virtual combinatorial libraries if we are to reliably rank compounds for testing. Alternatively, focusing predictions on compounds that have a higher maximal similarity value (above a set threshold) to compounds in the training set is also justified to improve predictions. Future studies should also attempt to compare descriptors and build computational models from screening datasets for Mtb grown under different conditions that may be physiologically relevant and mimic the heterogeneity in the disease conditions.

Understanding the quality of the compounds and avoiding those with liabilities such as reactivity is also important if we are to take some compounds beyond preclinical tests. Our results with SMARTS filtering of the Novartis data are similar to those observed with other screening datasets (12) and should be considered before selecting compounds for further testing. The number of alerts failures also correlates with the number of Lipinski violations (Fig. 4A). Hence, the majority of compounds with Lipinski violations ≥ 2 (which represents approximately 11% of FDA approved drugs, Fig. 4B) fails the SMARTS filters from different groups. Although several groups have been moving away from using the Lipinski Rule of Five (22) as a hard cut-off for molecule selection in recent years, and towards other kinds of filters, it would appear that compounds failing such SMARTS filters are also likely to have higher numbers of Lipinski violations, which is a novel finding. This perhaps further indicates how general the Lipinski rules are and how they may still be useful as an indicator in filtering compounds for screening, whether for neglected or other diseases. In general, the TB community is not alone in its search for filtering methods to insure the highest quality molecules are selected for follow-up hit-to-lead optimization.

Antituberculosis drug discovery could use these Bayesian models (or other computational methods reviewed elsewhere (19)) and SMARTS alerts to assist in selecting compounds for *in vitro* screening that may have a higher probability of activity against Mtb while at the same time a lower probability of undesirable off-target effects due to chemical reactivity. In the words of Yogi Berra, "Life is a learning experience, only if you learn." We need to learn from the very large whole cell TB screening datasets that are now in the public domain in order to expedite the discovery of useful chemical probes of Mtb that could in turn lead to novel therapeutic treatments.

ACKNOWLEDGMENTS

REFERENCES

1. Balganesh TS, Alzari PM, Cole ST. Rising standards for tuberculosis drug development. Trends Pharmacol Sci. 2008;29:576–81.
2. Zhang Y. The magic bullets and tuberculosis drug targets. Annu Rev Pharmacol Toxicol. 2005;45:529–64.
3. Ballel L, Field RA, Duncan K, Young RJ. New small-molecule synthetic antimycobacterials. Antimicrob Agents Chemother. 2005;49:2153–63.
4. Schneider G. Virtual screening: an endless staircase? Nat Rev Drug Discov. 2010;9:273–6.
5. Prathipati P, Ma NL, Keller TH. Global Bayesian models for the prioritization of antitubercular agents. J Chem Inf Model. 2008;48:2362–70.
6. Cho Y, Ioerger TR, Sacchettini JC. Discovery of novel nitro-benzothiazole inhibitors for Mycobacterium tuberculosis ATP phosphoribosyl transferase (HisG) through virtual screening. J Med Chem. 2008;51:5984–92.
7. Oprea TI, Davis AM, Teague SJ, Leeson PD. Is there a difference between leads and drugs? A historical perspective. J Chem Inf Comput Sci. 2001;41:1308–15.
8. Jones DR, Ekins S, Li L, Hall SD. Computational approaches that predict metabolic intermediate complex formation with CYP3A4 (+b5). Drug Metab Dispos. 2007;35:1466–75.
9. Ananthan S, Faaleolea ER, Goldman RC, Hobrath JV, Kwong CD, Laughon BE, *et al.* High-throughput screening for inhibitors of Mycobacterium tuberculosis H37Rv. Tuberculosis (Edinburgh, Scotland). 2009;89:334–53.
10. Ekins S, Kaneko T, Lipinksi CA, Bradford J, Dole K, Spektor A, *et al.* Analysis and hit filtering of a very large library of compounds screened against Mycobacterium tuberculosis. Mol Biosyst. 2010;6:2316–24.
11. Ekins S, Williams AJ. Meta-analysis of molecular property patterns and filtering of public datasets of antimalarial "hits" and drugs. Med Chem Comm. 2010;1:325–30.
12. Ekins S, Williams AJ. When pharmaceutical companies publish large datasets: an abundance of riches or fool's gold? Drug Disc Today. 2010;15:812–5.
13. Huth JR, Mendoza R, Olejniczak ET, Johnson RW, Cothron DA, Liu Y, *et al.* ALARM NMR: a rapid and robust experimental method to detect reactive false positives in biochemical screens. J Am Chem Soc. 2005;127:217–24.
14. Huth JR, Song D, Mendoza RR, Black-Schaefer CL, Mack JC, Dorwin SA, *et al.* Toxicological evaluation of thiol-reactive compounds identified using a la assay to detect reactive molecules by nuclear magnetic resonance. Chem Res Toxicol. 2007;20:1752–9.

15. Metz JT, Huth JR, Hajduk PJ. Enhancement of chemical rules for predicting compound reactivity towards protein thiol groups. J Comput Aided Mol Des. 2007;21:139–44.

16. Williams AJ, Tkachenko V, Lipinski C, Tropsha A, Ekins S. Free online resources enabling crowdsourced drug discovery. Drug Discov World. 2009;10:33–8.

17. Lougheed KE, Taylor DL, Osborne SA, Bryans JS, Buxton RS. New anti-tuberculosis agents amongst known drugs. Tuberculosis (Edinburgh, Scotland). 2009;89:364–70.

18. Lamichhane G, Freundlich JS, Ekins S, Wickramaratne N, Bishai WR. Essential metabolites of M. tuberculosis and their Mimics. Mbio. 2011;2:e00301–10.

19. Ekins S, Freundlich JS, Choi I, Sarker M, Talcott C. Computational databases, pathway and cheminformatics tools for tuberculosis drug discovery. Trends Microbiol. 2011;19:65–74.

20. Ekins S, Bradford J, Dole K, Spektor A, Gregory K, Blondeau D, et al. A collaborative database and computational models for tuberculosis drug discovery. Mol Biol Syst. 2010;6:840–51.

21. http://www.collaborativedrug.com/register. http://www.collaborativedrug.com/register.

22. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 1997;23:3–25.

23. Zientek M, Stoner C, Ayscue R, Klug-McLeod J, Jiang Y, West M, et al. Integrated in silico-in vitro strategy for addressing cytochrome P450 3A4 time-dependent inhibition. Chem Res Toxicol. 2010;23:664–76.

24. Ghose AK, Viswanadhan VN, Wendoloski JJ. Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: an analysis of ALOGP and CLOGP methods. J Phys Chem. 1998;102:3762–72.

25. Langdon SR, Mulgrew J, Paolini GV, van Hoorn WP. Predicting cytotoxicity from heterogeneous data sources with Bayesian learning. J Chem Inf. 2010;2:11.

26. Hann M, Hudson B, Lewell X, Lifely R, Miller L, Ramsden N. Strategic pooling of compounds for high-throughput screening. J Chem Inf Comput Sci. 1999;39:897–902.

27. Blake JF. Identification and evaluation of molecular properties related to preclinical optimization and clinical fate. Med Chem (Shariqah (United Arab Emirates)). 2005;1:649–55.

28. Stover CK, Warrener P, VanDevanter DR, Sherman DR, Arain TM, Langhorne MH, et al. A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. Nature. 2000;405:962–6.

29. Matsumoto M, Hashizume H, Tomishige T, Kawasaki M, Tsubouchi H, Sasaki H, et al. OPC-67683, a nitro-dihydro-imidazooxazole derivative with promising action against tuberculosis in vitro and in mice. PLoS Med. 2006;3:e466.

30. Walters WP, Ajay, Murcko MA. Recognizing molecules with drug-like properties. Curr Opin Chem Biol. 1999;3:384–7.

31. Gopalakrishnan B, Aparna V, Jeevan J, Ravi M, Desiraju GR. A virtual screening approach for thymidine monophosphate kinase inhibitors as antitubercular agents based on docking and pharmacophore models. J Chem Inf Model. 2005;45:1101–8.

32. Maddry JA, Ananthan S, Goldman RC, Hobrath JV, Kwong CD, Maddox C, et al. Antituberculosis activity of the molecular libraries screening center network library. Tuberculosis (Edinburgh, Scotland). 2009;89:354–63.